JOURNAL OF APPLIED STATISTICS
2021, VOL. 48, NO. 8, 1496–1512
https://doi.org/10.1080/02664763.2020.1789077

Taylor & Francis
Taylor & Francis Group

Maximum likelihood estimation for the proportional odds model with mixed interval-censored failure time data

Liang Zhu[a], Xingwei Tong[b], Dingjiao Cai[b], Yimei Li[c], Ryan Sun[d], Deo K. Srivastava[c] and Melissa M. Hudson[e]

[a]CCTS, University of Texas Health Science Center at Houston, Houston, TX, USA; [b]School of Mathematical Sciences, Beijing Normal University, Beijing, People's Republic of China; [c]Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA; [d]Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [e]Departments of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN, USA

ABSTRACT

This article discusses regression analysis of mixed interval-censored failure time data. Such data frequently occur across a variety of settings, including clinical trials, epidemiologic investigations, and many other biomedical studies with a follow-up component. For example, mixed failure times are commonly found in the two largest studies of long-term survivorship after childhood cancer, the datasets that motivated this work. However, most existing methods for failure time data consider only right-censored or only interval-censored failure times, not the more general case where times may be mixed. Additionally, among regression models developed for mixed interval-censored failure times, the proportional hazards formulation is generally assumed. It is well-known that the proportional hazards model may be inappropriate in certain situations, and alternatives are needed to analyze mixed failure time data in such cases. To fill this need, we develop a maximum likelihood estimation procedure for the proportional odds regression model with mixed interval-censored data. We show that the resulting estimators are consistent and asymptotically Gaussian. An extensive simulation study is performed to assess the finite-sample properties of the method, and this investigation indicates that the proposed method works well for many practical situations. We then apply our approach to examine the impact of age at cranial radiation therapy on risk of growth hormone deficiency in long-term survivors of childhood cancer.

ARTICLE HISTORY
Received 13 May 2020
Accepted 22 June 2020

KEYWORDS
Maximum likelihood estimation; mixed interval-censored data; Proportional odds model

1. Introduction

Interval-censored failure time data arise when a failure time of interest is not observed exactly but is only known to lie within an interval [12]. This type of data appears across many different research fields, including epidemiologic investigations, medical studies, and social science experiments. An extensive body of literature is available for the statistical analysis of interval-censored failure time data [4,19]. However, less work has focused on

CONTACT Ryan Sun rsun3@mdanderson.org

the case of mixed interval-censored data, which occurs when a dataset consists of a mixture of exact and interval-censored failure times.

Mixed interval-censored data occur very commonly in real-world situations, often due to ubiquitous data challenges such as missing data or integration of multiple data resources. For example, interval-censored outcomes arise frequently in the Childhood Cancer Survivor Study (CCSS), one of the largest datasets available for studying the late effects of childhood cancer treatment. This ongoing multi-institutional study follows over 30,000 childhood cancer survivors and 5000 healthy siblings by distributing periodic questionnaires that ask participants whether a number of specific medical conditions have occurred. If the condition has occurred, subjects are further asked the age of first occurrence. A single follow-up questionnaire can contain 140 such questions querying cardiovascular, respiratory, hormonal, urinary, digestive, and neurological issues, among other conditions. The time to event outcomes are more difficult to collect than the binary indicators of event occurrence for a variety of reasons, for example, because subjects have difficulty remembering exact dates. Thus some event occurrences are not paired with an event time, and the missing event time can only be inferred to fall between the cancer diagnosis and the time of a first follow-up or between two consecutively recorded follow-ups, resulting in an interval-censored outcome. The rate of missingness in event time over a single 140-item CCSS questionnaire can rise as high as 55%. Therefore, the time-to-event data from an given outcome consists of a mixture of interval-censored event times and exact event times.

Datasets integrated from multiple resources may also result in mixed interval-censored data. Such an issue arises in the the St. Jude Lifetime Cohort (SJLIFE), another large survivorship study that follows a cohort of over 5,000 childhood cancer survivors who were treated for their cancer at St. Jude Childrens Research Hospital (St. Jude). Late effects related to cancer treatments are recorded through two different modalities in SJLIFE - ascertained retrospectively via St. Jude medical records or identified prospectively during SJLIFE visits. For events that are recorded in the medical records, exact event times are available. For diseases newly detected during SJLIFE visits, the onset time is unknown but can be inferred to fall between the last and current visits, resulting in interval-censored data.

Mixed interval-censored data arise frequently in other biomedical studies that require periodic follow-up, especially for chronic diseases. Depending on the monitoring schedule and recollection of subjects, age at event onset can generally be known for some patients and only inferred to fall within an interval for others. Additional prominent examples include the Framingham Heart Disease Study [15], the Danish Diabetes Study [18], and the Sudan HIV/AIDS Study[7].

Although mixed interval-censored data arise frequently in practice, there exists only limited literature on their analysis. Huang [10] considered the nonparametric maximum likelihood estimation (MLE) of a distribution function based on mixed interval-censored data; Zhao et al. [24,25] developed some generalized log-rank tests for the nonparametric comparison of survival functions. For regression analysis specifically, the only existing reports appear to be [13] and [8], which investigated the fitting of the proportional hazards model and a parametric Cox model, respectively. It is well known that the proportional hazards model may not be appropriate in some situations, and alternative regression models for mixed interval-censored data are needed for such scenarios.

Unlike the proportional hazards model, the proportional odds model specifies that covariates have multiplicative effects on the odds function rather than on the hazard function. Under the proportional odds model, the hazard ratio approaches unity as time increases, so it is usually preferred when the covariate effects diminish over time. Another situation in which the proportional odds model is often preferred is when there are many ties in the data [3]. The proportional odds model has been discussed extensively [1,11,14,17,22] for the case of right-censored or interval-censored data, but an established approach for mixed interval-censored data does not yet exist. In this work, we develop the MLE approach for fitting a proportional odds model to mixed interval-censored data.

The remaining sections of the article are organized as follows. In Section 2, we introduce some notation and describe the proportional odds model as well as the structure of mixed interval-censored data. In Section 3, we develop an MLE procedure for fitting the proportional odds model to mixed interval-censored data. In particular, we show that the resulting estimators of both regression parameters and the baseline distribution function are asymptotically consistent and Gaussian under some mild regularity conditions. In Section 4, we present the results obtained from an extensive simulation study conducted to assess the finite-sample performance of the proposed method. These simulations suggest that the approach works well for many practical situations. In Section 6, we apply the approach to the data arising from the SJLIFE study described above, and in Section 6 we present a discussion and concluding remarks.

2. Notation, models, and likelihood function

Consider a failure time study that consists of n subjects. For subject i, let T_i denote the failure time of interest, and suppose that there exists a $d-$dimensional vector of covariates denoted by Z_i. Also suppose that given Z_i, T_i follows the proportional odds model

$$\text{logit}F_T(t|Z_i) = \text{logit}F_0(t) + Z_i'\beta_0. \tag{1}$$

Here $F_T(t|Z_i)$ denotes the conditional distribution function of T given Z_i, F_0 denotes a baseline distribution function, β_0 denotes a vector of regression parameters, and $\text{logit}(x) = \log x/(1-x)$. Under the above model, it is easy to see that the conditional survival density functions, given Z_i, have the forms

$$S_T(t|Z_i) = \frac{1}{1 + H_0(t)\exp(Z_i'\beta_0)}$$

and

$$f_T(t|Z_i) = \frac{\exp(Z_i'\beta_0)}{[1 + H_0(t)\exp(Z_i'\beta_0)]^2} h_0(t),$$

respectively. In the above, $H_0(t) = F_0(t)/(1-F_0(t))$, which is assumed to be a strictly increasing function with continuous positive derivative function $h_0(t) = \mathrm{d}H_0(t)/\,\mathrm{d}t$.

In the following, we will assume that there exist two potential examination times for each subject, denoted by U_i and V_i with $U_i \le V_i$. The event time is left-censored if $T_i \le U_i$, interval-censored if $U_i < T_i \le V_i$, or right-censored if $T_i > V_i$. Also, it will be assumed

that given Z_i, T_i is independent of (U_i, V_i), and the joint distribution of (U_i, V_i, Z_i) does not involve β_0 and H_0. Then the likelihood function of $\theta = (\beta, H)$ can be written as

$$L_n(\theta) = \prod_{i=1}^{n} \left[\frac{\exp(Z_i'\beta)}{[1 + H(T_i)\exp(Z_i'\beta)]^2} h(T_i) \right]^{\delta_{Ei}} \left[1 - \frac{1}{1 + H(U_i)\exp(Z_i'\beta)} \right]^{\delta_{Li}} \times \left[\frac{1}{1 + H(U_i)\exp(Z_i'\beta)} - \frac{1}{1 + H(V_i)\exp(Z_i'\beta)} \right]^{\delta_{Ii}} \left[\frac{1}{1 + H(V_i)\exp(Z_i'\beta)} \right]^{\delta_{Ri}},$$

where δ_{Ei}, δ_{Li}, δ_{Ii}, and δ_{Ri} are indicator functions that take the value 1 if T_i is observed exactly, left-censored, interval-censored, or right-censored, respectively. Thus $\delta_{Ei} + \delta_{Li} + \delta_{Ii} + \delta_{Ri} = 1$. Define $H\{t\} = H(t) - H(t-)$ to be the size of the jump in H at time point t. Then the log-likelihood function of θ has the form

$$\begin{aligned} l_n(\theta) = {} & \sum_{i=1}^{n} \delta_{Ei} \left[Z_i'\beta - 2\log(1 + H(T_i)\exp(Z_i'\beta)) + \log(H\{T_i\}) \right] \\ & + \delta_{Li} \log \left[1 - \frac{1}{1 + H(U_i)\exp(Z_i'\beta)} \right] + \delta_{Ri} \log \left[\frac{1}{1 + H(V_i)\exp(Z_i'\beta)} \right] \\ & + \delta_{Ii} \log \left[\frac{1}{1 + H(U_i)\exp(Z_i'\beta)} - \frac{1}{1 + H(V_i)\exp(Z_i'\beta)} \right]. \end{aligned}$$

3. Maximum likelihood estimation

In this section, we discuss the MLE of the parameter θ based on the log-likelihood function $l_n(\theta)$. Let $0 = T_{(0)} < T_{(1)} < \cdots$ denote the distinct ordered times among all exactly observed T_i. Following [10], note that for any finite sample size n, $F_0(t)$, and hence $H_0(t)$ is determined in the likelihood above only at the exactly observed failure times and at the examination times (U_i, V_i) for $i = 1, \ldots, n$. For subject i, define $\delta_i = 1$ if the exact failure time is available for subject i and $\delta_i = 0$ otherwise. Then the log-likelihood function of θ can be rewritten as

$$\begin{aligned} l_n(\theta) &= \sum_{i=1}^{n_1} \log[S_T(T_{(i)}|Z_i) - S_T(T_{(i-1)})|Z_i)] + \sum_{i=1}^{n_2} \log[S_T(U_i|Z_i) - S_T(V_i|Z_i)] \\ &= \sum_{i=1}^{n_1} \log\left[\{1 + H(T_{(i-1)})\exp(Z_i'\beta)\}^{-1} - \{1 + H(T_{(i)})\exp(Z_i'\beta)\}^{-1} \right] \\ &\quad + \sum_{i=1}^{n_2} \log\left[\{1 + H(U_i)\exp(Z_i'\beta)\}^{-1} - \{1 + H(V_i)\exp(Z_i'\beta)\}^{-1} \right], \end{aligned}$$

where $n_1 = \sum_{i=1}^{n} \delta_i$, and $n_2 = n - n_1$. Let $t_1 < \cdots < t_M$ denote the M distinct times among all observed T_i, U_i, and V_i, and define $H = (H(t_1), \ldots, H(t_M))$. It is apparent that for the maximization of $l_n(\theta)$, one needs to focus only on the values of H at the t_m, $m = 1, \ldots, M$, and consider the step functions that jump only at the t_m. Also, let $(L_i, R_i]$ denote the observed interval that contains the time to event for the ith subject.

For subject i, define $\alpha_{im} = 1$ if $(L_i, R_i]$ contains t_m and $\alpha_{im} = 0$ otherwise. Also, define $h_m = H(t_m) - H(t_{m-1})$ and $\gamma_m = \log h_m$. It follows that

$$l_n(\beta, \gamma) = \sum_{i=1}^{n} \log \left[\alpha_{i1} + \sum_{m=1}^{M} (\alpha_{im+1} - \alpha_{im}) \{1 + H(t_m) \exp(Z_i' \beta)\}^{-1} \right]$$
$$= \sum_{i=1}^{n} \log \left[\alpha_{i1} + \sum_{m=1}^{M} (\alpha_{im+1} - \alpha_{im}) \{1 + \sum_{j=1}^{m} \exp(\gamma_j + Z_i' \beta)\}^{-1} \right],$$

where $\gamma = (\gamma_1, \ldots, \gamma_M)'$. Thus one can maximize the log-likelihood function above by using, for example, the Newton-Raphson algorithm. To maximize $l_n(\beta, \gamma)$ using Newton-Raphson, we need to calculate the first and second derivatives of $l_n(\beta, \gamma)$. The detailed formulas are provided in the Appendix.

Let $S_0(t)$ denote the true baseline survival function for the failure time T. Let $\hat{\beta}_n$ and $\hat{H}_n$ denote the maximum-likelihood estimators of β and H_0 defined above and take $\hat{S}_n(t) = 1/(1 + \hat{H}_n(t))$, the maximum-likelihood estimator of the baseline survival function. Also let τ denote the longest follow-up time and $S_0(t)$ denote the true baseline survival function of the T_i. The two theorems below give the asymptotic properties of $\hat{\beta}_n$ and $\hat{S}_n$. Together with Lemma 1 in the Appendix, we can obtain the identifiability of the parameters for this model and describe the following asymptotic behavior of the proposed maximum likelihood estimator.

Theorem 3.1: *Suppose that conditions* C1–C5 *given in the Appendix hold. Then as* $n \to \infty$, *we have*

$$\hat{\beta}_n \to \beta_0 \text{ a.s.} \quad \text{and} \quad \sup_{t \in [0,\tau]} |\hat{S}_n(t) - S_0(t)| \to 0 \text{ a.s.}$$

Theorem 3.2: *Suppose that conditions* C1–C5 *given in the Appendix hold. Then as* $n \to \infty$, *we have*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \to N(0, \Sigma), \quad \sqrt{n}(\hat{S}_n - S_0) \to W,$$

where Σ *denotes the asymptotic covariance matrix of* $\hat{\beta}_n$ *and* W *denotes a zero-mean Gaussian process.*

The proof of the results given above is sketched in the Appendix. For inference about β_0, one needs to estimate Σ. One way to do this is to use the the Fisher information matrix $I(\hat{\beta}, \hat{\gamma})$, which appears to be reasonable based on the numerical results below.

4. Simulation

In this section, we present some results obtained from an extensive simulation study conducted to assess the finite-sample performance of the method proposed above. We considered a situation with one covariate Z_i generated from the uniform distribution over $(0, 1)$. We then used model (1) to generate failure times under Case 1: $H_0(t) = t$ and Case 2: $H_0(t) = (1 + t/2)^2 - 1$. For the observation times, we generated the first time point O_{1i} from the uniform distribution over $\{0, 1, 2, \ldots, \tau - 1\}$ and then the second time point O_{2i}

Table 1. Estimation of regression parameters with case 1 for $n = 200$.

		Proposed method				exact T only				Interval-censored T only				Peto's method			
β	p	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP
−1	0.25	−1.04	0.53	0.54	0.95	−1.06	0.88	0.85	0.97	−1.04	0.67	0.71	0.93	−1.03	0.53	0.54	0.95
	0.5	−1.02	0.48	0.51	0.93	−1.02	0.61	0.62	0.94	−1.04	0.82	0.88	0.93	−1.02	0.48	0.51	0.93
	0.75	−1.03	0.46	0.46	0.95	−1.02	0.50	0.49	0.97	−1.12	1.52	1.36	0.95	−1.03	0.46	0.46	0.95
−0.5	0.25	−0.52	0.54	0.55	0.95	−0.56	0.88	0.84	0.97	−0.50	0.70	0.74	0.94	−0.52	0.54	0.55	0.95
	0.5	−0.52	0.49	0.49	0.95	−0.53	0.61	0.60	0.96	−0.51	0.87	0.93	0.95	−0.52	0.49	0.49	0.95
	0.75	−0.52	0.46	0.45	0.96	−0.52	0.50	0.48	0.96	−0.53	1.60	1.49	0.95	−0.52	0.46	0.45	0.96
0	0.25	−0.01	0.56	0.56	0.95	−0.06	0.87	0.83	0.97	0.03	0.75	0.77	0.94	−0.01	0.56	0.56	0.95
	0.5	−0.02	0.50	0.50	0.96	−0.04	0.61	0.59	0.96	0.00	0.94	0.98	0.95	−0.02	0.50	0.50	0.96
	0.75	−0.01	0.46	0.45	0.96	−0.02	0.49	0.48	0.97	0.03	1.73	1.66	0.96	−0.01	0.46	0.45	0.96
0.5	0.25	0.50	0.59	0.58	0.95	0.44	0.87	0.83	0.97	0.56	0.81	0.82	0.95	0.50	0.58	0.58	0.95
	0.5	0.49	0.52	0.51	0.95	0.47	0.61	0.60	0.97	0.54	1.01	1.07	0.96	0.49	0.51	0.51	0.95
	0.75	0.49	0.47	0.46	0.95	0.48	0.50	0.48	0.96	0.59	1.80	1.83	0.96	0.49	0.47	0.46	0.95
1	0.25	1.01	0.62	0.61	0.96	0.94	0.88	0.83	0.97	1.10	0.89	0.88	0.96	1.01	0.61	0.60	0.96
	0.5	1.00	0.53	0.53	0.96	0.97	0.62	0.60	0.96	1.09	1.12	1.16	0.96	1.00	0.53	0.53	0.96
	0.75	1.00	0.48	0.47	0.95	0.98	0.50	0.48	0.97	1.30	3.57	3.87	0.97	0.99	0.47	0.47	0.95

from the uniform distribution over $\{ O_{1i} + 1, \ldots, \tau \}$ with $\tau = 10$. We generate the indicator δ_i from a Bernoulli distribution with success probability p. If $\delta_i = 1$ then the exact failure time T_i is recorded. Otherwise, an interval is recorded, and if the failure time T_i is less than O_{1i}, then the interval is recorded as $(0, O_{1i}]$. If the failure time T_i is between O_{1i} and O_{2i}, then the interval is recorded as $(O_{1i}, O_{2i}]$. If the failure time T_i is greater than O_{2i}, then the interval is recorded as (O_{2i}, ∞). The results given below are based on 1000 replications with the sample size $n = 200$ or 400.

Table 1 shows the results of the estimation of the regression parameter β_0 for Case 1 with $n = 200$, $\beta_0 = -1, -0.5, 0, 0.5$, or 1, and $p = 0.25$, 0.5, or 0.75. The table includes the average of point estimates $\hat{\beta}_n$ (AVE), the average of the estimated standard errors of $\hat{\beta}_n$ (SEE), the sample standard deviation of $\hat{\beta}_n$ (SSE), and the 95% empirical coverage probability (CP). For comparison, we also obtained the estimates that would result from fitting proportional odds model to only the observed exact failure times or only the observed interval-censored times. This simplified alternative discards much of the data.

One can see that the proposed estimator appears to be unbiased and that the variance estimation and normal distribution approximations are reasonable. Furthermore, the proposed method is clearly more efficient than the method that makes use of only exactly observed or only interval-censored data. The results of $n = 400$ for Case 1 and the results of $n = 200$ and $n = 400$ for Case 2 are presented in Tables 2–4 and offer similar conclusions.

Another simplified approach for mixed interval-censored data is to transfer each exact failure time to an interval-censored observation by adding a small amount of time on either side of the observed point as proposed by Peto & Peto [16]. We also applied this approach in the simulation study and fitted the proportional odds model to the transformed interval-censored data. The results are included in Tables 1–4. In Figure 1, we plotted the averages of the proposed estimates $\hat{H}_n$ and the pointwise confidence bands for Case 2 with $n = 400$, $p = 0.75$ and $\beta = 0.5$ and 1. We can see that the simplified method and the proposed method provide similar parameter estimates, but the confidence bands of $\hat{H}_n$ are wider when using the simplified approach, indicating an efficiency loss.

Table 2. Estimation of regression parameters with Case 1 for $n = 400$.

		Proposed method				exact T only				Interval-censored T only				Peto's method			
β	p	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP
−1	0.25	−1.01	0.37	0.36	0.96	−1.00	0.62	0.62	0.95	−1.03	0.47	0.46	0.96	−1.01	0.37	0.36	0.96
	0.5	−1.01	0.34	0.33	0.96	−1.00	0.43	0.42	0.95	−1.04	0.58	0.58	0.96	−1.01	0.34	0.33	0.96
	0.75	−1.01	0.32	0.32	0.96	−1.01	0.35	0.35	0.95	−1.07	0.84	0.85	0.96	−1.01	0.32	0.32	0.96
−0.5	0.25	−0.51	0.38	0.37	0.96	−0.50	0.61	0.61	0.95	−0.52	0.49	0.49	0.96	−0.51	0.38	0.37	0.96
	0.5	−0.51	0.35	0.34	0.96	−0.50	0.43	0.42	0.95	−0.52	0.60	0.60	0.96	−0.50	0.35	0.34	0.96
	0.75	−0.51	0.32	0.32	0.96	−0.51	0.35	0.35	0.95	−0.54	0.88	0.90	0.96	−0.51	0.32	0.32	0.96
0	0.25	−0.01	0.39	0.39	0.96	0.00	0.61	0.61	0.95	−0.02	0.52	0.52	0.96	−0.01	0.39	0.39	0.96
	0.5	0.00	0.24	0.28	0.65	0.01	0.29	0.35	0.65	−0.02	0.43	0.53	0.94	0.00	0.24	0.28	0.65
	0.75	−0.01	0.32	0.32	0.96	−0.01	0.35	0.35	0.95	−0.03	0.94	0.98	0.95	−0.01	0.32	0.32	0.95
0.5	0.25	0.50	0.41	0.41	0.95	0.50	0.61	0.61	0.95	0.50	0.56	0.55	0.96	0.50	0.41	0.41	0.95
	0.5	0.50	0.36	0.36	0.95	0.50	0.43	0.43	0.96	0.50	0.69	0.69	0.96	0.50	0.36	0.36	0.95
	0.75	0.48	0.33	0.33	0.95	0.48	0.35	0.36	0.94	0.48	1.02	1.08	0.95	0.48	0.33	0.33	0.95
1	0.25	1.00	0.43	0.43	0.95	1.00	0.61	0.62	0.95	1.00	0.61	0.60	0.96	1.00	0.43	0.43	0.95
	0.5	1.00	0.37	0.37	0.95	0.99	0.43	0.43	0.96	1.01	0.76	0.76	0.96	1.00	0.37	0.37	0.95
	0.75	1.00	0.33	0.34	0.95	0.99	0.35	0.36	0.95	1.06	1.13	1.26	0.94	0.99	0.33	0.34	0.95

Table 3. Estimation of regression parameters with Case 2 for $n = 200$.

		Proposed method				exact T only				Interval-censored T only				Peto's method			
β	p	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP
−1	0.25	−1.03	0.62	0.60	0.97	−1.06	0.88	0.84	0.97	−0.99	0.87	0.87	0.95	−1.02	0.62	0.60	0.97
	0.50	−1.03	0.54	0.52	0.97	−1.03	0.62	0.60	0.96	−1.02	1.12	1.08	0.95	−1.03	0.54	0.52	0.96
	0.75	−1.03	0.49	0.46	0.96	−1.03	0.50	0.49	0.96	−1.01	1.77	1.70	0.94	−1.03	0.49	0.46	0.96
−0.5	0.25	−0.52	0.62	0.61	0.97	−0.56	0.87	0.83	0.97	−0.49	0.91	0.90	0.96	−0.52	0.62	0.60	0.97
	0.50	−0.53	0.53	0.52	0.96	−0.53	0.61	0.60	0.96	−0.53	1.20	1.13	0.97	−0.53	0.53	0.52	0.96
	0.75	−0.52	0.48	0.46	0.96	−0.52	0.50	0.48	0.96	−0.51	2.03	2.02	0.95	−0.52	0.47	0.46	0.96
0	0.25	0.00	0.64	0.62	0.96	−0.06	0.87	0.83	0.97	0.06	0.97	0.97	0.96	0.00	0.63	0.62	0.96
	0.50	−0.02	0.54	0.53	0.96	−0.03	0.61	0.59	0.96	0.05	1.26	1.23	0.97	−0.01	0.54	0.53	0.96
	0.75	−0.01	0.47	0.47	0.96	−0.02	0.49	0.48	0.97	0.09	2.13	2.84	0.93	−0.01	0.47	0.47	0.96
0.5	0.25	0.50	0.66	0.64	0.97	0.44	0.87	0.83	0.97	0.59	1.07	1.07	0.96	0.50	0.66	0.64	0.96
	0.50	0.49	0.55	0.54	0.96	0.47	0.61	0.59	0.97	0.57	1.40	1.38	0.97	0.49	0.55	0.54	0.96
	0.75	0.49	0.48	0.47	0.96	0.48	0.50	0.48	0.96	0.69	2.43	3.70	0.93	0.49	0.48	0.47	0.96
1	0.25	1.01	0.71	0.68	0.97	0.94	0.88	0.83	0.97	1.13	1.19	1.24	0.97	1.01	0.70	0.68	0.97
	0.50	1.00	0.57	0.56	0.96	0.97	0.62	0.60	0.97	1.10	1.56	1.59	0.96	1.00	0.57	0.56	0.96
	0.75	1.00	0.49	0.48	0.96	0.98	0.50	0.48	0.96	1.95	3.18	11.71	0.92	1.00	0.49	0.48	0.96

5. Application to SJLIFE

We applied the proposed method to a subset of the data from the SJLIFE study. One of the primary objectives of the SJLIFE study is to evaluate endocrine-related late effects such as growth hormone deficiency (GHD) in patients with brain irradiation. GHD is associated with neuropsychiatric cognitive, cardiovascular, neuromuscular, metabolic, and skeletal abnormalities, and it may increase the risk of premature death. In addition, GHD is the hormone deficiency most likely to go undiagnosed. To obtain information about each participant's GHD, researchers extracted the histories of pituitary deficiencies from subjects' medical records if available. For those who had no medical record of GHD occurrence, a fasting blood sample was collected at the time of the SJLIFE evaluation to investigate the presence of pituitary deficiency. More details on the study can be found in [2].

Table 4. Estimation of regression parameters with Case 2 for $n = 400$.

		Proposed method				exact T only				Interval-censored T only				Peto's method			
β	p	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP	AVE	SEE	SSE	CP
−1	0.25	−1.02	0.42	0.42	0.96	−1.00	0.61	0.62	0.96	−1.04	0.58	0.57	0.96	−1.02	0.42	0.42	0.96
	0.5	−1.02	0.37	0.36	0.96	−1.00	0.43	0.42	0.95	−1.06	0.72	0.72	0.96	−1.01	0.37	0.36	0.96
	0.75	−1.02	0.34	0.33	0.96	−1.01	0.35	0.35	0.95	−1.13	1.11	1.09	0.94	−1.02	0.33	0.33	0.96
−0.5	0.25	−0.51	0.43	0.43	0.95	−0.50	0.61	0.61	0.95	−0.52	0.61	0.62	0.95	−0.51	0.43	0.43	0.95
	0.5	−0.51	0.37	0.37	0.95	−0.50	0.43	0.42	0.95	−0.53	0.76	0.77	0.95	−0.51	0.37	0.37	0.96
	0.75	−0.51	0.33	0.33	0.96	−0.51	0.35	0.35	0.95	−0.58	1.16	1.20	0.95	−0.51	0.33	0.33	0.96
0	0.25	−0.01	0.44	0.45	0.95	0.00	0.61	0.61	0.95	−0.01	0.65	0.66	0.96	−0.01	0.44	0.44	0.95
	0.5	−0.01	0.38	0.38	0.95	0.00	0.43	0.42	0.96	−0.02	0.82	0.83	0.96	−0.01	0.44	0.44	0.95
	0.75	−0.01	0.33	0.34	0.95	−0.01	0.35	0.35	0.95	−0.05	1.25	1.33	0.95	−0.01	0.44	0.44	0.95
0.5	0.25	0.49	0.46	0.47	0.96	0.50	0.61	0.61	0.95	0.50	0.71	0.72	0.95	0.49	0.46	0.47	0.96
	0.5	0.49	0.38	0.38	0.95	0.50	0.43	0.43	0.96	0.50	0.89	0.91	0.96	0.49	0.38	0.38	0.96
	0.75	0.49	0.34	0.34	0.95	0.49	0.35	0.35	0.95	0.53	1.38	1.55	0.95	0.49	0.34	0.34	0.95
1	0.25	1.01	0.48	0.49	0.96	0.99	0.61	0.62	0.95	1.03	0.79	0.82	0.96	1.00	0.48	0.49	0.96
	0.5	1.00	0.40	0.39	0.96	0.99	0.43	0.43	0.95	1.04	1.02	1.03	0.96	1.00	0.39	0.39	0.96
	0.75	1.00	0.34	0.35	0.95	0.99	0.35	0.36	0.95	1.15	1.53	1.89	0.93	1.00	0.34	0.34	0.95

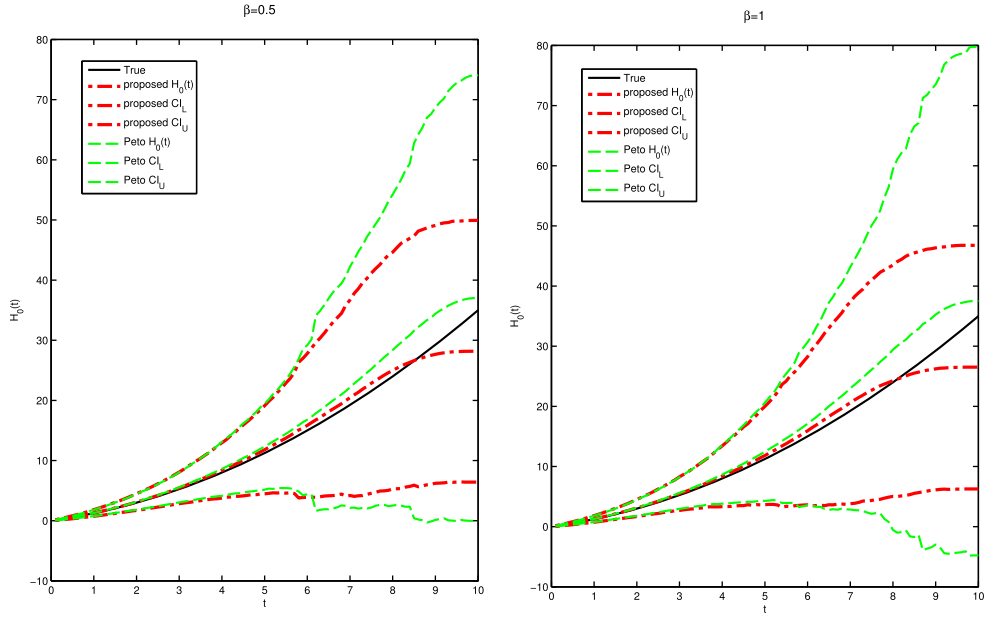

Figure 1. Estimates of the baseline function $H_0(t) = (1 + t/2)^2 - 1$.

In our work, we focused on a subgroup of participants who were younger than 10 years old when they received cranial radiation therapy (CRT), as this population was found to be at high-risk of GHD [2]. Our objective was to further investigate whether the risk of the first occurrence of GHD depended on the age at CRT, i.e. whether the risk of GHD was higher in the childhood cancer survivors who received CRT before 5 years old compared to those survivors who received CRT when they were at least 5 years old. In total there are 524 participants in the subgroup, and 269 experienced GHD. Among those with GHD, 113 (42%) had GHD diagnoses in their medical record and hence provided an exact failure

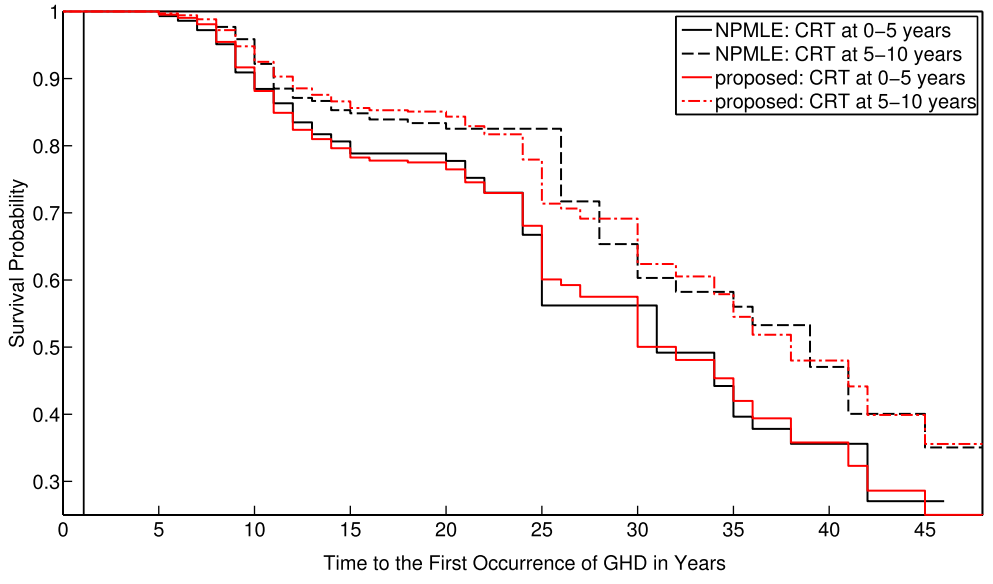

Figure 2. Estimates of the baseline survival function $S_0(t)$.

time. One hundred and fifty-six (58%) survivors had GHD identified at SJLIFE evaluations, and so the onset time was only known to fall in an interval between two St. Jude visits. For the analysis, we defined $Z_i = 0$ if the participant received CRT before 5 years old (n = 301) and $Z_i = 1$ (n = 223) otherwise. The application of the proposed method produced a $\hat{\beta}_n$ of -0.504 with the estimated standard error of 0.172. This corresponds to a p-value of 0.0033 for testing no difference in the risks of the first occurrence of GHD between the two age groups. GHD appears to occur significantly less in the survivors who received CRT at an older age. For comparison, using only the interval-censored data produced an estimate of -0.473 (standard error 0.216), using only the exact failure time data produced an estimate of -0.553 (0.232), and treating exact failure times as interval-censored observations produced an estimate of -0.502 (0.194). All estimated standard errors with alternative approaches were larger than the estimated standard error using the proposed method.

Figure 2 presents the estimated survival functions $\hat{S}_n$ given by the proposed method for the two age groups. We obtained the nonparametric estimates of the two survival functions given by [10] as well. These results suggest that the proportional odds model provides a good fit, as the proposed and nonparametric estimates are close to each other.

6. Discussion and concluding remarks

In this paper, we discussed the regression analysis of mixed interval-censored data arising from the proportional odds model. As detailed previously, such data arise quite frequently in medical studies with a follow-up component. We developed a maximum likelihood estimation approach and established both the finite-sample and asymptotic properties of the resulting estimators. The numerical results suggest that the proposed method works well for practical situations and better than some simplified alternatives. We also applied the

method to analyze the impact of age at CRT treatment on GHD risk in childhood cancer survivors. Code for implementation of the ideas or the simulation is available upon request.

Future research on this problem can go in several directions. First, note that for each subject, it was assumed that there exists two random observation time points. Such data are often referred to as Case-2 interval-censored data [19]. A more general situation corresponding to this is the so-called case-K interval-censored data, where there exists K observation points or a sequence of observation time points. It would be useful to generalize the proposed method to case-K interval-censored data. In addition, the focus of this paper is the proportional odds model. Sometimes neither the proportional odds model or the proportional hazards model fits the data well; thus, one may consider other models, such as the additive hazards model or linear transformation model. Model checking is another possible direction for future research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was partly supported by NIH [R03DE029238] to Zhu and Li and also supported by National Science Funding of China 11971064 to Dr. Tong.

References

[1] S. Bennett, *Analysis of survival data by the proportional odds model*, Stat. Med. 2 (1983), pp. 273–277.
[2] W. Chemaitilly, Z. Li, S. Huang, K.K. Ness, K.L. Clark, D.M. Green, N. Barnes, G.T. Armstrong, M.J. Krasin, D.K. Srivastava, C.H. Pui, T.E. Merchant, L.E. Kun, A. Gajjar, M.M. Hudson, L.L. Robison, and C.A. Sklar, *Anterior hypopituitarism in adult survivors of childhood cancers treated with cranial radiotherapy: a report from the st Jude lifetime cohort study*, J. Clin. Oncol. 33 (2015), pp. 492–500.
[3] S. Chen and A.K. Manatunga, *A note on proportional hazards and proportional odds models*, Stat. Probab. Lett. 77 (2007), pp. 981–988.
[4] D. Chen, J. Sun, and K. Peace, *Interval-Censored Time-to-Event Data: Methods and Applications*, CRC Press, Boca Raton, FL, 2012.
[5] K. Chen and X. Tong, *Varying coefficient transformation models with censored data*, Biometrika 97 (2010), pp. 969–976.
[6] M. Chen, X. Tong, and J. Sun, *The proportional odds model for multivariate interval-censored failure time data*, Stat. Med. 26 (2007), pp. 5147–5161.
[7] V.G. De Gruttola and S.W. Lagakos, *Analysis of doubly-censored survival data, with application to AIDS*, Biometrics 45 (1989), pp. 1–11.
[8] F. Elfaki, M. Azram, and M. Usman, *Parametric Cox's model for partly interval-censored data with applications to AIDS studies*, Int. J. Appl. Phys. Math. 2 (2012), pp. 352–354.
[9] J. Huang, *Efficient estimation for the proportional hazards model with interval censoring*, Ann. Stat. 24 (1996), pp. 540–568.
[10] J. Huang, *Asymptotic properties of nonparametric estimation based on partly interval-censored data*, Stat. Sin. 9 (1999), pp. 501–519.
[11] J. Huang and A.J. Rossini, *Sieve estimation for the proportional-odds failure-time regression model with interval censoring*, J. Am. Stat. Assoc. 92 (1997), pp. 960–967.
[12] J.D. Kalbfleisch and R.L. Prentice, *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, Hoboken, NJ, 2011.

[13] J.S. Kim, *Maximum likelihood estimation for the proportional hazards model with partly interval-censored data*, J. Royal Stat. Soc.: Ser. B (Stat. Methodol.) 65 (2003), pp. 489–502.
[14] S.A. Murphy, A.J. Rossini, and A.W. Van Der Vaart, *Maximum likelihood estimation in the proportional odds model*, J. Am. Stat. Assoc. 92 (1997), pp. 968–976.
[15] P.M. Odell, K.M. Anderson, and R.B. D'Agostino, *Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model*, Biometrics 48 (1992), pp. 951–959.
[16] R. Peto and J. Peto, *Asymptotically efficient rank invariant test procedures*, J. Royal Stat. Soc. Ser. A (General) 135 (1972), pp. 185–207.
[17] D. Rabinowitz, R.A. Betensky, and A.A. Tsiatis, *Using conditional logistic regression to fit proportional odds models to interval censored data*, Biometrics 56 (2000), pp. 511–518.
[18] H.H. Ramlau, P. Kragh Andersen, K. Borch-Johnsen, and T. Deckerf, *Life insurance for insulin-dependent diabetics*, Scand. Actuar. J. 1987 (1987), pp. 19–36.
[19] J. Sun, *The Statistical Analysis of Interval-Censored Failure Time Data*, Springer, New York, NY, 2007.
[20] A.W. Van der Vaart and J.A. Wellner, *Existence and consistency of maximum likelihood in upgraded mixture models*, J. Multivar. Anal. 43 (1992), pp. 133–146.
[21] A.W. Van Der Vaart and J.A. Wellner, *Weak Convergence and Empirical Processes*, Springer, New York, NY, 1996.
[22] S. Yang and R.L. Prentice, *Semiparametric inference in the proportional odds regression model*, J. Am. Stat. Assoc. 94 (1999), pp. 125–136.
[23] D. Zeng, D. Lin, and G. Yin, *Maximum likelihood estimation for the proportional odds model with random effects*, J. Am. Stat. Assoc. 100 (2005), pp. 470–483.
[24] Q. Zhao and J. Sun, *Generalized log-rank test for mixed interval-censored failure time data*, Stat. Med. 23 (2004), pp. 1621–1629.
[25] X. Zhao, Q. Zhao, J. Sun, and J. Kim, *Generalized log-Rank tests for partly interval-Censored failure time data*, Biometrical J. 50 (2008), pp. 375–385.

Appendix

A.1 Appendix 1. The first and second derivatives of $l_n(\beta, \gamma)$.

To maximize $l_n(\beta, \gamma)$, we need to calculate the first and second derivatives of $l_n(\beta, \gamma)$. For this, following [6], let $b_{im} = \sum_{j=1}^{m} \exp(\gamma_j + Z_i'\beta)$ and

$$w_i = \alpha_{i1} + \sum_{m=1}^{M} (\alpha_{im+1} - \alpha_{im})(1 + b_{im})^{-1}.$$

Then we have

$$U_\beta(\beta, \gamma) = \frac{\partial l_n(\beta, \gamma)}{\partial \beta} = -\sum_{i=1}^{n} w_i^{-1} V_{\beta,i},$$

and

$$U_{\gamma_m}(\beta, \gamma) = \frac{\partial l_n(\beta, \gamma)}{\partial \gamma_m} = -\sum_{i=1}^{n} w_i^{-1} V_{\gamma_m,i},$$

where

$$V_{\beta,i} = Z_i \sum_{m=1}^{M} (\alpha_{im+1} - \alpha_{im})\, b_{im}\, (1 + b_{im})^{-2},$$

and

$$V_{\gamma_m,i} = \sum_{s=m}^{M} (\alpha_{is+1} - \alpha_{is})\, \exp(\gamma_m + Z_i\beta)\, (1 + b_{im})^{-2}.$$

Furthermore, one can determine that

$$D_{\beta\beta}(\beta,\gamma)=\frac{\partial^2 l_n(\beta,\gamma)}{\partial\beta\partial\beta'}=-\sum_{i=1}^{n}(w_i^{-2}\,V_{\beta,i}\,V'_{\beta,i}-w_i^{-1}\,V_{\beta\beta})\,,$$

$$D_{\beta\gamma_m}(\beta,\gamma)=\frac{\partial^2 l_n(\beta,\gamma)}{\partial\beta\partial\gamma_m}=-\sum_{i=1}^{n}(w_i^{-2}\,V_{\beta,i}\,V'_{\gamma_m,i}-w_i^{-1}\,V_{\beta\gamma_m})\,,$$

and

$$D_{\gamma_m\gamma_{m'}}(\beta,\gamma)=\frac{\partial^2 l_n(\beta,\gamma)}{\partial\gamma_m\partial\gamma_{m'}}=-\sum_{i=1}^{n}(w_i^{-2}\,V_{\gamma_m,i}\,V'_{\gamma_{m'},i}-w_i^{-1}\,V_{\gamma_m\gamma_{m'}})\,,$$

where

$$V_{\beta\beta}=Z_i\,Z_i'\sum_{m=1}^{M}(\alpha_{im+1}-\alpha_{im})\,b_{im}\,(b_{im}-1)\,(1+b_{im})^{-3}\,,$$

$$V_{\beta\gamma_m}=Z_i\sum_{s=m}^{M}(\alpha_{is+1}-\alpha_{is})\,\exp(\gamma_m+Z_i\beta)\,(b_{is}-1)\,(1+b_{is})^{-3}\,,$$

$$V_{\gamma_m\gamma_{m'}}=\sum_{s=m}^{M}(\alpha_{is+1}-\alpha_{is})\,(1-b_{is})\,(1+b_{is})^{-3}\,\exp(\gamma_m+Z_i\beta),\quad \text{for } m=m'\,,$$

and

$$V_{\gamma_m\gamma_{m'}}=2\exp(\gamma_m+Z_i\beta)\,\exp(\gamma_{m'}+Z_i\beta)\sum_{s=m}^{M}(\alpha_{is+1}-\alpha_{is})\,(1+b_{is})^{-3},\quad \text{for } m\neq m'\,.$$

A.2 Appendix 2. Proofs of Theorems 3.1 and 3.2

In this section, we will sketch the proofs of Theorems 3.1 and 3.2. Here we need the following regularity conditions:

C1. Let $\mathcal{B}$ be a bounded closed subset of R^d and $\mathcal{S}$ a compact set. Suppose that the true value (β_0, S_0) is an interior point of $(\mathcal{B}\times\mathcal{S})$.

C2. Let n_1, n_2, n_3 and n_4 denote the number of subjects who have exactly observed, left-censored, interval-censored, or right-censored observations, respectively. Suppose that

$$\lim_{n\to\infty}\frac{n_k}{n}=\alpha_k>0\quad\text{and}\quad\sum_{k=1}^{4}\alpha_k=1.$$

C3. (3a.) Suppose that there exist $0<\tau_0<\tau_1<\infty$ such that $P[\tau_0\le U<V\le\tau_1]=1$ and $0<H_0(\tau_0)<H_0(\tau_1)<\infty$. (3b.) Suppose that there exists a positive number η_0 such that $P(V-U\ge\eta_0)=1$.

C4. The covariate Z is uniformly bounded, that is, there exists $z_0>0$ such that $P(\|Z\|\le z_0)=1$.

C5. If $g(U,V)+Z'\beta=0$ almost surely for some deterministic function g and vector β, then $g\equiv 0$ and $\beta=0$.

Note that Conditions C1, C3 and C4 are very commonly assumed in the interval-censored data literature. Condition C2 is not too strict because the fraction of subjects in each category data is usually positive if we are in the mixed data setting. Condition C5 is needed for the identifiability of the proportional odds model and the covariate Z. To prove the asymptotic properties, let $Pg(X)=\int g(x)dF(x)$ and $P_ng(X)=n^{-1}\sum_{i=1}^{n}g(X_i)$ for any random variable X with distribution function F and function g.

Let the log-likelihood function of one sample be denoted as

$$\begin{aligned} l(\theta) = {} & \delta_E \left[Z'\beta - 2\log(1 + H(T)\exp(Z'\beta)) + \log(H\{T\}) \right] \\ & + \delta_L \log\left[1 - \frac{1}{1 + H(U)\exp(Z'\beta)} \right] + \delta_R \log\left[\frac{1}{1 + H(V)\exp(Z'\beta)} \right] \\ & + \delta_I \log\left[\frac{1}{1 + H(U)\exp(Z'\beta)} - \frac{1}{1 + H(V)\exp(Z'\beta)} \right]. \end{aligned}$$

Then we have the following lemma.

Lemma A.1: *If $l(\theta) = l(\theta_0)$ a.s., then $\theta = \theta_0$, where $\theta = (\beta, H) \in \Theta$.*

Proof: Since $\delta_E + \delta_L + \delta_R + \delta_I = 1$, we only prove the result when either of the four indicators holds true. If $\delta_E = 1$, then for any $t \le \tau$,

$$\int_0^t \frac{\exp(Z'\beta)}{1 + H(s)\exp(Z'\beta)}\, \mathrm{d}H(s) = \int_0^t \frac{\exp(Z'\beta_0)}{1 + H_0(s)\exp(Z'\beta_0)}\, \mathrm{d}H_0(s),$$

and thus $\log H(t)/H_0(t) = Z'(\beta - \beta_0)$ a.s. and thus $\theta = \theta_0$ derived from condition C5.

Similar techniques can be applied to get the same result when $\delta_R = 1$, or $\delta_L = 1$ or $\delta_I = 1$. This completes the proof. ■

Proof of Theorem 3.1: The proof is adapted from [9,13,20]. First we'll show that $\hat{H}_n\{t\}$ cannot be ∞. Otherwise, if $\hat{H}_n\{T_i\} = \infty$, then the ith log-likelihood function

$$< O(1) - \log(\hat{H}_n(T_i)) \to -\infty.$$

Then we show that $\sup_n \hat{H}_n(\tau) < \infty$. Set $\theta_n^* = (\hat{\beta}_n, H_n^*)$, where $H_n^* = \hat{H}_n/\xi_n, \xi_n = \hat{H}_n(\tau)$. Then

$$\begin{aligned} \frac{1}{n}(l_n(\hat{\theta}_n) - l_n(\theta_n^*)) = {} & \frac{1}{n}\sum_{i=1}^{n} \Big\{ \delta_{Ei}[-2\log(1 + \xi_n H_n^*(T_i)\exp(Z_i'\beta)) + \log(\xi_n)] \\ & + 2\delta_{Ei}\log(1 + H_n^*(T_i)\exp(Z_i'\beta)) \\ & + \delta_{Ii}\log[(1 + \xi_n H_n^*(U_i)\exp(Z_i'\beta))^{-1} - (1 + \xi_n H_n^*(V_i)\exp(Z_i'\beta))^{-1}] \\ & - \delta_{Ii}\log[(1 + H_n^*(U_i)\exp(Z_i'\beta))^{-1} - (1 + H_n^*(V_i)\exp(Z_i'\beta))^{-1}] \\ & + \delta_{Ri}[\log(1 + \xi_n H_n^*(V_i)\exp(Z_i'\beta))^{-1} - \log(1 + H_n^*(V_i)\exp(Z_i'\beta))^{-1}] \\ & + \delta_{Li}\left[\log\frac{\xi_n H_n^*(U_i)\exp(Z_i'\beta)}{1 + \xi_n H_n^*(U_i)\exp(Z_i'\beta)} - \log\frac{H_n^*(U_i)\exp(Z_i'\beta)}{1 + H_n^*(U_i)\exp(Z_i'\beta)} \right] \Big\} \\ = {} & -\log\xi_n \frac{1}{n}\sum_{i=1}^{n}\delta_{Ei} - 2\frac{1}{n}\sum_{i=1}^{n}\delta_{Ei}\log(\xi_n^{-1} + H_n^*(T_i)\exp(Z_i'\beta)) \\ & + 2\frac{1}{n}\sum_{i=1}^{n}\delta_{Ei}\log(1 + H_n^*(T_i)\exp(Z_i'\beta)) \\ & + \frac{1}{n}\sum_{i=1}^{n}\delta_{Ii}\log[(1 + \xi_n H_n^*(U_i)\exp(Z_i'\beta))^{-1} - (1 + \xi_n H_n^*(V_i)\exp(Z_i'\beta))^{-1}] \\ & - \frac{1}{n}\sum_{i=1}^{n}\delta_{Ii}\log[(1 + H_n^*(U_i)\exp(Z_i'\beta))^{-1} - (1 + H_n^*(V_i)\exp(Z_i'\beta))^{-1}] \end{aligned}$$

$$+\frac{1}{n}\sum_{i=1}^{n}\delta_{Ri}[\log(1+\xi_n H_n^*(V_i)\exp(Z_i'\beta))^{-1}-\log(1+H_n^*(V_i)\exp(Z_i'\beta))^{-1}]$$

$$+\frac{1}{n}\sum_{i=1}^{n}\delta_{Li}\left[\log\frac{\xi_n H_n^*(U_i)\exp(Z_i'\beta)}{1+\xi_n H_n^*(U_i)\exp(Z_i'\beta)}-\log\frac{H_n^*(U_i)\exp(Z_i'\beta)}{1+H_n^*(U_i)\exp(Z_i'\beta)}\right].$$

It is obvious that the last six terms are uniformly upper bounded because H_n^* is uniformly bounded by 1 for sufficiently large n. Therefore, if $\xi_n \to \infty$, then $l_n(\hat{\theta}_n) - l_n(\theta_n^*) \to -\infty$ as n tends to infinity. This means that for sufficiently large n, $l_n(\hat{\theta}_n) - l_n(\theta_n^*) < 0$ with high probability, which contradicts the fact that $l_n(\hat{\theta}_n) - l_n(\theta_n^*) \geq 0$ because $\hat{\theta}_n$ maximizes $l_n(\theta)$.

Now we show the consistency. Let $(\hat{\beta}_n, \hat{H}_n)$ be the MLE of (β_0, H_0) based on the observed data, and let $(\tilde{\beta}_n, \tilde{H}_n)$ be the MLE of (β_0, H_0) based on the exact data and right-censored data, respectively. Define $\hat{S}_n = 1/(1+\hat{H}_n)$, and $\tilde{S}_n = 1/(1+\tilde{H}_n)$. Let $x = (\delta_L, \delta_I, u, v, z)$ and write

$$\hat{p}(x) = p(x; \hat{\beta}_n, \hat{S}_n), \quad \tilde{p}(x) = p(x; \tilde{\beta}_n, \tilde{S}_n),$$

where $p(x;\beta,S) = \{1-S(u|z)\}^{\delta_L}\{S(u|z)-S(v|z)\}^{\delta_I}$ with $S(\cdot|z) = 1/(1+He^{z'\beta})$, and $H = (1-S)/S$.

Let E be the expectation given $\delta_E = \delta_R = 0$ under (β_0, S_0). For any $(\beta, S) \neq (\beta_0, S_0)$, by condition(5) and Jensen's inequality, for every $\alpha \in (0,1)$,

$$E\log\left[1+\alpha\left\{\frac{p(X;\beta,S)}{p_0(X)}-1\right\}\right] < 0, \tag{A1}$$

where $p_0(X) = p(X;\beta_0,S_0)$. Note that $\tilde{p}(x) \to p_0(x)$ for almost all $x = (\delta_L, \delta_I, u, v, z)$ by the consistency of $(\tilde{\beta}_n, \tilde{S}_n)$. See [14]. For an open ball $\mathcal{N}$ around (β, S), define

$$\tilde{p}(x;\mathcal{N}) = \sup_{(\beta',S')\in\mathcal{N}} p(x;\beta',S').$$

Then for a sequence of open balls $\mathcal{N}_\epsilon$ with radius ϵ shrinking to (β, S) as $\epsilon \to 0$, we have $\tilde{p}(x;\mathcal{N}_\epsilon) \to p(x;\beta,S)$. By (3), for ϵ sufficiently small and $N_1 = n_1 + n_2$ sufficiently large, there exists a $\eta_\epsilon > 0$ so that

$$E\log\left[1+\alpha\left\{\frac{\tilde{p}(X;\mathcal{N}_\epsilon)}{\tilde{p}(X)}-1\right\}\right] \wedge \eta_\epsilon < 0. \tag{A2}$$

From the definition of $(\hat{\theta}_n, \hat{S}_n)$ and $(\tilde{\theta}_{N_1}, \tilde{S}_{N_1})$, after simple algebra, we have

$$\prod_{j=1}^{N_2}\hat{p}(x_j) \geq \prod_{j=1}^{N_2}\tilde{p}(x_j),$$

where $N_2 = n - N_1$. By the concavity of the function $u \to \log u$, this implies

$$\sum_{j=1}^{N_2}\log\left[1+\alpha\left\{\frac{\hat{p}(x_j)}{\tilde{p}(x_j)}-1\right\}\right] \geq 0. \tag{A3}$$

For any open neighborhood $\mathcal{N}_0$ of the true value (β_0, S_0), its complement in $\mathcal{B}\times\mathcal{S}$ is a closed subset of a compact set, hence also compact. The open cover $\{\mathcal{N}_{(\beta,S)}, (\beta,S) \notin \mathcal{N}_0\}$ of this complement has a finite subcover $\{\mathcal{N}_{(\beta_1,S_1)}, \ldots, \mathcal{N}_{(\beta_m,S_m)}\}$. If $(\hat{\beta}_n, \hat{S}_n)$ is not in $\mathcal{N}_0$, it must be in one of the subcovers, in which case $\tilde{p}(x;\mathcal{N}_{(\beta_k,S_k)}) \geq \hat{p}(x)$ for every x. Thus by (5),

$$\{(\hat{\beta}_n,\hat{S}_n) \notin \mathcal{N}_0\} \subset \bigcup_{k=1}^{m}\left[\frac{1}{N_2}\sum_{j=1}^{N_2}\log\left[1+\alpha\left\{\frac{\tilde{p}(X_j;\mathcal{N}_{(\beta_k,S_k)})}{\tilde{p}(X_j)}-1\right\}\right] \wedge \eta_{(\theta_k,S_k)} \geq 0\right].$$

The probability of each of the sets in the union is the probability that an average of uniformly bounded and independent random variables is nonnegative. However these random variables have

negative expectation by (4). By Hoeffding's inequality , each of the probabilities is of order $e^{-\epsilon N_2}$, where ϵ can be chosen equal to $2\rho^2/(M - \log(1-\alpha))^2$. Here $M = \max(\eta_{(\beta_k,S_k)}, 1 \le k \le m)$, and ρ is any negative number that the expectation in (4) is less than ρ, say $N_1 \ge N$, where N is a sufficiently large integer. Consequently,

$$\sum_{N_2=1}^{\infty} \sup_{N_1 \ge N} P((\hat{\beta}_n,\ \hat{S}_n) \notin \mathcal{N}_0) < \infty.$$

By a minor modification of the Borel-Cantelli lemma and condition(2), it follows that, with probability 1, $(\hat{\beta}_n, \hat{S}_n) \in \mathcal{N}_0$ eventually. By the definition of our product topology, this implies that $\hat{\beta}_n \to \beta_0$ a.s., and $\hat{S}_n$ converges to S_0 almost surely under $P_{(\beta_0,S_0)}$, namely

$$\lim_{n\to\infty} \sup_{t\in[0,\tau]} |\hat{S}_n(t) - S_0(t)| = 0$$

almost surely. This implies, by the fact that $S = 1/(1+H)$,

$$\lim_{n\to\infty} \sup_{t\in[0,\tau]} |\hat{H}_n(t) - H_0(t)| = 0$$

almost surely. This completes the proof. ■

Proof of Theorem 3.2: The proof is adapted from [5]. We check the conditions of theorem 3.3.1 of [21]. Define a random map, ψ, as

$$\psi(\theta)[b,g] = \frac{\partial}{\partial a} f\left(\beta + ab, H + a\int g\,\mathrm{d}H\right)\Big)\Big|_{a=0},$$

where f is the logarithm of the likelihood function of $\theta = (\beta, H)$ based on one sample $(T, Z, U, V, \delta_E, \delta_R, \delta_L, \delta_I)$, specifically,

$$\begin{aligned}
f(\theta) = f(\beta,H) &= \delta_E\left[Z'\beta - 2\log(1 + H(T)\exp(Z'\beta)) + \log(H\{T\})\right] \\
&\quad + \delta_R \log\left[\frac{1}{1+H(V)\exp(Z'\beta)}\right] + \delta_L \log\left[1 - \frac{1}{1+H(U)\exp(Z'\beta)}\right] \\
&\quad + \delta_I \log\left[\frac{1}{1+H(U)\exp(Z'\beta)} - \frac{1}{1+H(V)\exp(Z'\beta)}\right].
\end{aligned}$$

$$\begin{aligned}
\psi(\theta)[b,g] &= \delta_E\left\{\frac{[1-H(T)\exp(Z'\beta)]Z'b}{1+H(T)\exp(Z'\beta)} - \frac{2\int_0^T g\,\mathrm{d}H_0\exp(Z'\beta)}{1+H(T)\exp(Z'\beta)} + g(T)\right\} \\
&\quad - \delta_R\left\{\frac{\int_0^V g\,\mathrm{d}H\exp(Z'\beta) + H(V)\exp(Z'\beta)Z'b}{1+H(V)\exp(Z'\beta)}\right\} \\
&\quad + \delta_L\left\{\frac{\int_0^U g\,\mathrm{d}H\exp(Z'\beta) + H(U)\exp(Z'\beta)Z'b}{H(U)\exp(Z'\beta)(1+H(U)\exp(Z'\beta))}\right\} \\
&\quad + \delta_I\left\{\frac{[1-H(U)H(V)\exp(2Z'\beta)]Z'b}{(1+H(U)\exp(Z'\beta))(1+H(V)\exp(Z'\beta))}\right. \\
&\quad + \frac{(1+H(V)\exp(Z'\beta))\int_0^U g\,\mathrm{d}H}{(H(V)-H(U))(1+H(U)\exp(Z'\beta))} \\
&\quad \left. - \frac{(1+H(U)\exp(Z'\beta))\int_0^V g\,\mathrm{d}H}{(H(V)-H(U))(1+H(V)\exp(Z'\beta))}\right\}.
\end{aligned}$$

Let $\Psi_n(\theta)[b,g] = P_n\psi(\theta)[b,g]$ and $\Psi(\theta)[b,g] = P\psi(\theta)[b,g]$. It is easy to prove that $\psi(\theta)[b,g]$ is Lipschitz in θ. Following lemma 7.1 of [9], one can prove that

$$\{\psi(\theta)[b,g] : \|b\| \le 1,\ g \in \mathcal{G}, \|\beta - \beta_0\|_2 + \|H - H_0\|_2 < \delta\}$$

is P_0 Donsker for some $\delta > 0$ and

$$\sup_{\|b\|\le 1, g\in\mathcal{G}} P_0(\psi(\theta)[b,g] - \psi(\theta_0)[b,g])^2 \to 0 \quad \text{as } \theta \to \theta_0.$$

Hence, condition(3.3.2) of theorem 3.3.1 of [21] is satisfied.

Now we prove that $\dot{\Psi}(\theta_0)$ is continuously invertable. Following the proof of Theorem 2 in [23], we only need to prove that, if $\psi(\theta_0)[b,g] = 0$ almost surely, then $b = 0, g = 0$. By letting $\delta_E = 1$, we obtain that

$$\frac{[1 - H_0(T)\exp(Z'\beta_0)]Z'b}{1 + H_0(T)\exp(Z'\beta_0)} - \frac{2\int_0^T g\,\mathrm{d}H_0(T)\exp(Z'\beta_0)}{1 + H_0(T)\exp(Z'\beta_0)} + g(T) = 0.$$

Let $\epsilon = H_0(T)\exp(Z'\beta_0), a = \exp(-Z'\beta_0), g = gH_0^{-1}$, then we have

$$\frac{1-\epsilon}{1+\epsilon}Z'b - \frac{2\int_0^T g(t)\,\mathrm{d}H_0(t)\exp(Z'\beta_0)}{1+\epsilon} + g(T) = 0.$$

Thus,

$$(1-\epsilon)Z'b - 2\int_0^{\epsilon} g(sa)\,\mathrm{d}s + g(\epsilon a)(1+\epsilon) = 0.$$

By taking the second derivative with respect to ϵ, we get

$$a^2 g''(\epsilon a)(1+\epsilon) = 0,$$

hence $g''(\epsilon a) = 0$. Let $g(x) = c_1 x + c_2$, we have

$$(1-\epsilon)Z'b - 2\int_0^{\epsilon}(c_1 sa + c_2)\,\mathrm{d}s + (c_1 a\epsilon + c_2)(1+\epsilon) = 0,$$

namely

$$(1-\epsilon)Z'b - c_2\epsilon + c_1 a\epsilon + c_2 = 0.$$

Thus

$$-Z'b - c_2 + c_1 a = 0$$
$$Z'b + c_2 = 0.$$

Consequently, $b = 0, c_1 = 0$, and $c_2 = 0$, which means $b = 0, g = 0$. When $\delta_R = 1$, we obtain

$$-\frac{1}{[1 + H_0(V)\exp(Z'\beta_0)]}\left[\int_0^V g\,\mathrm{d}H_0\exp(Z'\beta_0) + H_0(V)\exp(Z'\beta_0)Z'b\right] = 0.$$

Thus $H_0^{-1}(V)\int_0^V g\,\mathrm{d}H_0 + Z'b = 0$. It then follows from condition 5 that $b = 0$ and $g = 0$. When $\delta_L = 1$, we have

$$\int_0^U g\,\mathrm{d}H_0\exp(Z'\beta_0) + H_0(U)\exp(Z'\beta_0)Z'b = 0.$$

Thus $H_0^{-1}(U)\int_0^U g\,\mathrm{d}H_0 + Z'b = 0$. From condition 5, we get $b = 0$ and $g = 0$. Finally, when $\delta_I = 1$, we have

$$\left[H_0(U) + H_0(U)H_0^2(V)\exp(2Z'\beta_0) - H_0(V) - H_0(V)H_0^2(U)\exp(2Z'\beta_0)\right]Z'b$$
$$+ (1 + H_0(V)\exp(Z'\beta_0))^2\int_0^U g\,\mathrm{d}H_0 - (1 + H_0(U)\exp(Z'\beta_0))^2\int_0^V g\,\mathrm{d}H_0 = 0.$$

By taking the derivative with respect to U and V successively, we obtain that

$$(2H_0(V)\exp(Z'\beta_0) - H_0(U)\exp(Z'\beta_0))Z'b + g(U)[1 + H_0(V)\exp(Z'\beta_0)] - g(V)[1 + H_0(U)\exp(Z'\beta_0)] = 0.$$

By the symmetry of U and V, we also have

$$(2H_0(U)\exp(Z'\beta_0) - H_0(V)\exp(Z'\beta_0))Z'b + g(V)[1 + H_0(U)\exp(Z'\beta_0)] - g(U)[1 + H_0(V)\exp(Z'\beta_0)] = 0.$$

Then it is easy to see that

$$[H_0(U) + H_0(V)]Z'b = 0,$$

thus $b = 0$. Therefore,

$$(1 + H_0(V)\exp(Z'\beta_0))^2 \int_0^U g\,\mathrm{d}H_0 - (1 + H_0(U)\exp(Z'\beta_0))^2 \int_0^V g\,\mathrm{d}H_0 = 0.$$

By taking the derivative with respect to U and V successively, we have

$$[g(U)H_0(V) - g(V)H_0(U)]\exp(Z'\beta_0) + g(U) - g(V) = 0,$$

which implies that $g = 0$.

By similar arguments to [11], one can show that $\Psi_n(\hat{\theta}_n)[b, g] = o_p(n^{-1/2})$ and thus the fourth condition is satisfied since $\Psi(\theta_0)[b, g] = 0$. This completes the proof. ■